

# Making Pacific Language Materials Discoverable

Andrea Berez-Kroeker

Danielle Yarbrough

Ryan Shelby

*University of Hawai'i at Mānoa*

*Department of Linguistics*

Eleanor Kleiber

Mike Chohey

*University of Hawai'i at Mānoa*

*Library*

West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# Our presentation in a nutshell



About the Pacific Collection at UHM Hamilton Library

Languages of the Pacific

Problems in the Collection catalog - language materials could not be found!

Our project to fix these problems

West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# About the UHM Library Pacific Collection

World-premier collection of Pacific-related materials

100,000+ books, periodicals, microfilm, rare items

Policy to collect as much as possible in or about Pacific languages and linguistics >>

Most comprehensive collection of items in the world in or about Pacific languages

(cf Australian National University, the National Libraries of Australia and New Zealand, the US Library of Congress, and the University of the South Pacific)



West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



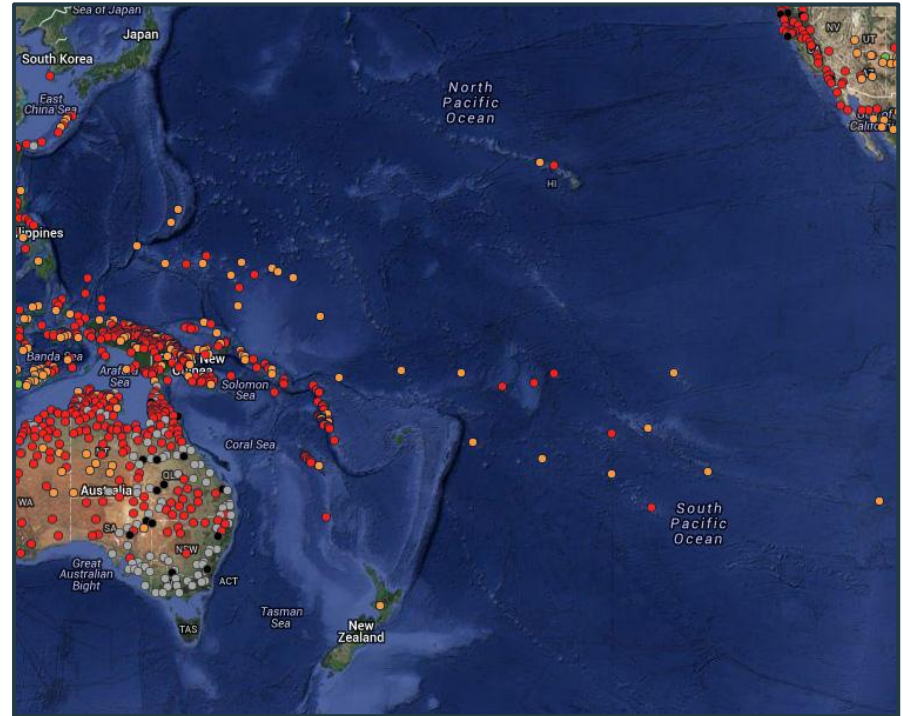
# Languages in the Pacific

Among the most diverse regions in the world!

At least 1,400 distinct languages

Thousands of dialects

Library users need to be able to find materials in catalog based on language



West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# The problem

Libraries *do* catalog books by language...but not very well at all!

Two ways:

1. MARC language codes: **515** 3-letter codes for the world's languages
  - a. Problem! There are **nearly 7,000** languages in the world
  - b. Problem! MARC uses “aggregate codes” for vast swaths of the world:

*map* “Austronesian other”

*paa* “Papuan other”

*cpe* “Creoles & Pidgins, English-based”

Hundreds of languages “included” here, no way to find them in catalog

West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# The problem

Libraries *do* catalog books by language...but not very well at all!

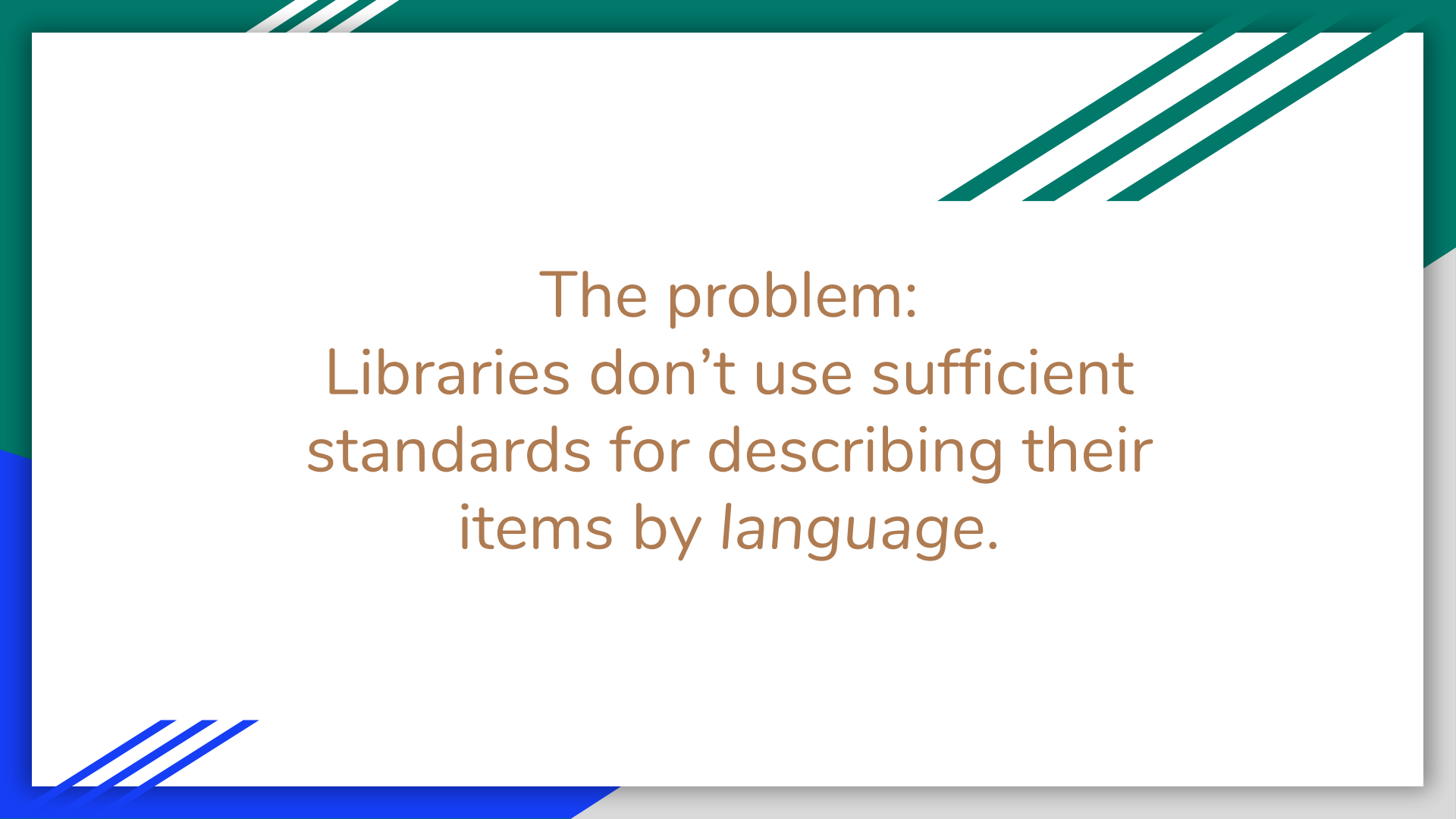
Two ways:

2. Library of Congress Subject Headings: A descriptive phrase, but 50% of Pacific languages *do not have* a LCSH

Example: Pááfang language materials have LCSH

“Hall Islands (Micronesia) -- Languages”

Not possible to find Pááfang materials in the catalog.



The problem:  
Libraries don't use sufficient  
standards for describing their  
items by *language*.

# Our grant funded project

Fortunately there *is* a sufficient standard that lists all the world's languages:

## ISO 639-3:

International Organization for Standardization Codes for  
the representation of names of languages

- Maintained by language experts & internationally recognized
- Aims to cover all 7,000+ natural human languages
- Curated by The Ethnologue/SIL International ([www.ethnologue.com](http://www.ethnologue.com))

**Our NEH project aim: to add ISO 639-3 codes to the Pacific Collection catalog**

West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017





# More about “the ISO codes”

Unambiguously identifies languages

Alternate spellings:

wpb for 1 language alternately spelled *Warlpiri*, *Wailbri*, *Walbiri*, *Walpiri*

Alternate names:

wos for 1 language known as *Hanga Hundi*, *Kwasengen* and *West Wosera*

Identical names for distinct languages:

kue for *Kuman* language of Papua New Guinea


kdi for *Kuman* language of Uganda

# The Ethnologue

“Value-added” gallery of info about the languages that ISO 639-3 identifies.

Can be changed/updated by request from user community (you!).

Not perfect, but invaluable to our project.

**Ethnologue**  
Languages of the World

Login | Shopping Cart

Find a language or country...

Languages | Countries | Products | Subscribe | About

## Kuman

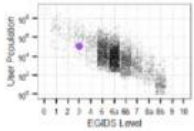
Print

LANGUAGE | MAP | FEEDBACK

### A language of Papua New Guinea

ISO 639-3	<a href="#">kue</a>
Alternate Names	Chimbu, Simbu
Population	115,000 (2000 census), increasing. Few monolinguals.
Location	Chimbu province: northern third; Eastern Highlands province: Daulo, Lufa and Unggal-Benna Lufa districts.
Language Maps	<a href="#">Papua New Guinea, Map 9</a> <a href="#">Papua New Guinea, Map 10</a>
Language Status	3 (Wider communication). Major area language used as a trade language for all the other Chimbu languages.
Classification	<a href="#">Trans-New Guinea, Chimbu-Wahgi, Chimbu</a>
Dialects	Kuman, Nagane (Genagane, Genogane), Yongomugl.
Typology	SOV; clause chaining, auxiliary verbs with adjuncts.
Language Use	Vigorous. All domains. Oral use in local administration, religion, local commerce. All ages. Positive attitudes. Most also use Tok Pisin [ <a href="#">tpi</a> ]. Some also use English [ <a href="#">eng</a> ]. Used as L2 by Chuave [ <a href="#">civ</a> ], Dom [ <a href="#">doa</a> ], Golin [ <a href="#">gvl</a> ], Nomane [ <a href="#">nof</a> ], Salt-Yui [ <a href="#">slj</a> ], Sinasina [ <a href="#">sst</a> ].
Language Development	Literacy rate in L1: 15%. Literacy rate in L2: 35%–40%. Used in first 3 grades in school. 20,000 can read and write Kuman. Taught in primary schools. Dictionary. Grammar. NT: 2008.
Language Resources	<a href="#">OLAC resources in and about Kuman</a>
Writing	Latin script [ <a href="#">Latn</a> ].

#### PLACE IN LANGUAGE CLOUD



Click to enlarge with explanation

#### ETHNOLOGUE PRODUCTS

Languages of Papua New Guinea, Map 9  
*An Ethnologue Language Map*  
**\$24.95**  
[Add to cart](#) [Browse all products](#)

#### JOIN THE CONVERSATION

[Create a free account](#) to post Feedback about a country or language and be notified when others do the same.

# Scope of our project

Our task:

To identify the **all the languages\*** found in the entire Pacific Collection

- All printed materials, including grey literature
- Main stacks and In-Cataloging
- Rare Book Vault
- Creole Collection (Pacific creoles only)

\* defined as *any amount of text in a Pacific language*, beyond a few token, well-known vocabulary items.

# Workflow step 1 - Compiling “the list”

Needed to identify which of the 100,000+ items would be part of the project.

Initial list: the ~9,550 items that already were cataloged for language with MARC or LCSH

Then added books in subjects that were likely to contain languages:

- Linguistics, Education, Anthropology, Biodiversity, works with titles in a Pacific language...

Final list: about 17,000 items to examine over three years.

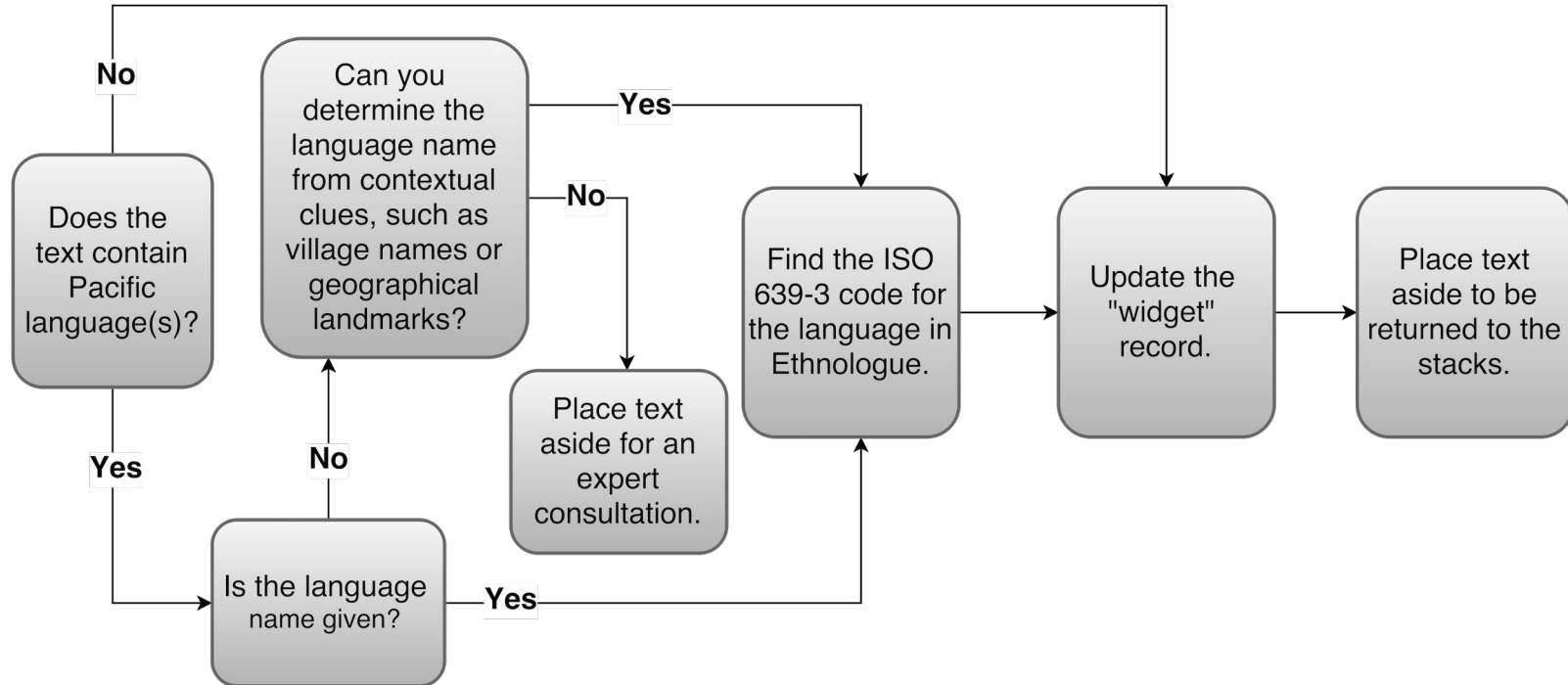


# Workflow step 2 - Prioritizing the items

The 17,000 items then sorted into three different priorities:

- **1st priority:** items already cataloged with MARC collective codes *map*, *paa*, or *cpe*.  
(These items definitely have language in them, but needed our eyes on them)
- **2nd priority:** books not yet cataloged for language.  
(These items likely to have language in them, but not “guaranteed”)
- **3rd priority:** books already coded with one-to-one MARC codes  
(These could be automatically updated in the library database by catalog specialist Chohey)

# Workflow step 3- Daily workflow



# Workflow step 4 - Increasing discoverability

Our final task: make our updated catalog information more findable by the public. Two ways:

- Added a **language-based search function** to the library catalog, allowing users to search by a language's ISO 639-3 code.
- Converted the new catalog information into a format that can be read by a **major language search engine** called OLAC (Open Language Archives Community)

[\(Try it here\)](#)

West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# Results

We anticipate that **11,500 items** will have been updated by this project.

- Resources are now **more discoverable by language**.

- Languages known by **multiple names are discoverable** under ISO 639-3.

- Showcases the **diversity of Pacific languages**

Cataloging by individual languages rather than by generic group identifiers.

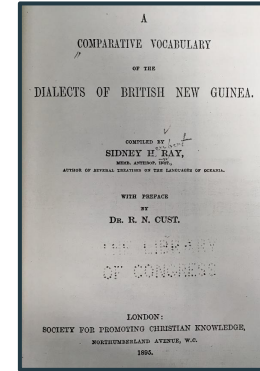
UHM Pacific Collection the first library area collection (that we know of) to include ISO 639-3 codes in its catalog.





# Example: MARC collective code was insufficient

*A comparative vocabulary of the dialects of British New Guinea*, by Sidney Herbert Ray



**Before:** Contains language data in more than 30 languages, but cataloged only with the MARC code *paa* ("Papuan other").

**Now:** Cataloged with 31 separate ISO 639-3 codes.

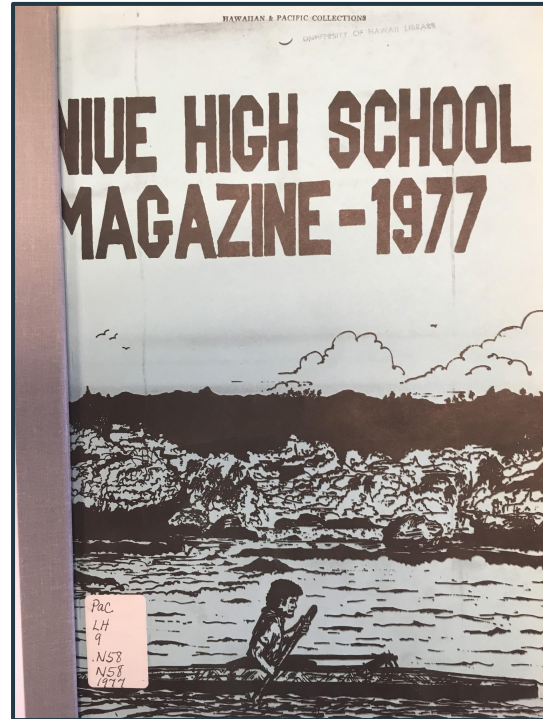
Libraries ISO 639-2	Linguists ISO 639-3
2% of languages have 1-to-1 codes	100% of languages have 1-to-1 codes
041 0_  a engpaa	041 _7  a eng  a mek  a nrz  a kbt  a don  a meu  a hul  a khz  a sbe  a swp  a tte  a tbo  a dob  a kij  a myw  a mpx  a tgo  a tof  a kjd  a ulk  a klq  a iar  a tqo  a oro  a kbk  a kqi  a bch  a huf  a jae  a ksd  a aia  a erk  2 iso639-3

## Example: Not previously coded for Pacific Language

Niue High School Magazine

**Before:** No language mentioned in catalog, despite containing considerable amounts of text in Niuean.

**Now:** Cataloged with Niuean's ISO 639-3 code niu.

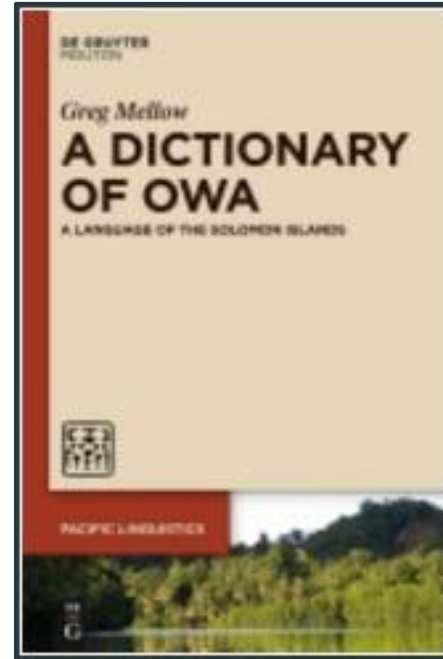


## Example: A language with alternative names

*A dictionary of Owa: A language of the Solomon Islands*, by Greg Mellow

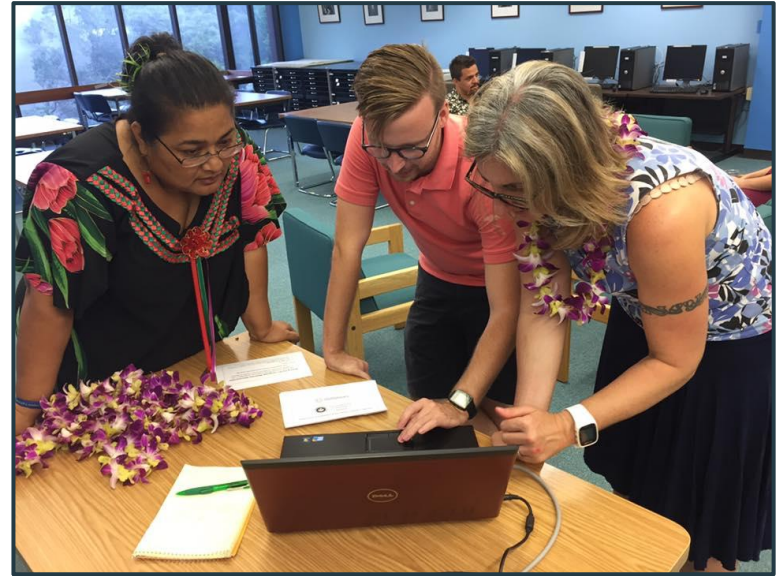
**Before:** Owa has five alternate names (Anganiwai, Anganiwei, Narihua, Santa Anna, Wanoni)! If users searched for any of them this item wouldn't have been in the result list.

**Now:** Library users can search the ISO 639-3 code `stn` and find this volume.



## Sharing with the community

On April 27th, 2017 we hosted a reception to demonstrate the project to the public.



West Maui Conference on Pacific Peoples & Their Environments, Lahaina HI October 13-14 2017



# Looking ahead

Project could easily be undertaken at other Pacific collections

- Australian National University
- the National Libraries of Australia and New Zealand
- the University of the South Pacific

Libraries with areal collections from other linguistically-diverse regions (southeast Asia; South America) could undertake similar projects.

Greater linguistic inclusion among library cataloging practices worldwide recognizes the importance of Indigenous voices.

# Mahalo nui loa!

Our hosts here in Maui.

Roxane Gaedeke, Carol Carl, Andrea Nakamura, Ann Rabinko, and Wes Polka.

Stu Dawrs, Dore Minatodani, Jodie Mattos, and Kapena Shim.

Ellie Seaton, Daniel Ishimitsu, and Nackil Sung.

Gary Simons, Bill Palmer, Åshild Naess, Alex François, Kaliko Trapp, Bob Blust, Piet Lincoln, and Jeff Siegel.

This work was supported in part by a grant from the National Endowment for the Humanities (PD-50034-14). All errors of omission and commission belong to the authors alone.